

DIVA: An Automatic Music Arrangement Technique Based on Impression of Images

Kisa OHYAMA and Takayuki ITOH

Ochanomizu University
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610 Japan
{kisa, itot}@itolab.is.ocha.ac.jp

Abstract. This poster reports our first approach for automatic music arrangement based on the impression of input images. Given a digital image and keywords of objects shot in the image, the technique selects a rhythm pattern associated from the keywords and color distribution of the image. As a preprocessing, the technique first provides sample colors, images, and keywords to users, and then collects the feedback of selection of rhythm patterns associated from them. The technique then leads equations to calculate the association of rhythm pattern from arbitrary input images. Finally, the technique automatically calculates the association scores of all prepared rhythm patterns from the images, and provides music arranged applying the associated rhythm pattern.

Key words: Automatic music arrangement, Digital image, Standard MIDI File(SMF), Color contradistinction

1 Introduction

Impression-based matching of music and image is a potential research topic. For example, several multimedia player software show arbitrary movies while playing the music, but users often complain the combination of music and movies. We think good matching of music and image will make multimedia software more fun. Recently many studies in this research area have been presented, such as music search engine based on impression of scenes, correlation of music and images via impression-related words, icon selection based on properties of music, and some sonification works; however, we think there are several more applications to be discussed.

This poster presents DIVA (Digital Image Varies Arrangement), a technique for automatic music arrangement based on impression of images. Input and output of DIVA are defined as follows:

Input: A standard MIDI File(SMF) consists of four parts, including melody, harmony, bass, and percussions. A raster image with some keywords of objects shot in the image.

Output: An arranged SMF whose melody and harmony are same as the input SMF, but bass and percussions are arranged based on the impression of the input image.

DIVA consists of preparation and actual stages. Preparation stage provides sample colors, images, keywords, and music arranged applying the prepared rhythm patterns, to users. It collects the selection of rhythm patterns from the users, as their feedback. It then leads equations to calculate association of music arrangement, from colors and keywords of arbitrary input images. Actual stage calculates associations of rhythm patterns from a given input image, and finally arranges the input music applying the most associated rhythm pattern.

2 Implementation

DIVA posterizes images into predefined 66 colors, and then leads equations to calculate association of rhythm patterns from colors. Our implementation defines the 66 colors as follows. It divides HSB color space into 63 subspaces, by dividing Hue into 7, Saturation into 3, and Brightness into 3, and then selects one color in the each subspace. Moreover, our definition includes 3 monochrome colors (white, gray, and black). Experimentally we found that the definition was good, since more colors did not bring more satisfaction of users, but fewer colors might decrease their satisfaction.

DIVA then calculates importance of i -th color C_i in an image, for each of 66 colors. Current implementation considers of color contradistinction to calculate C_i , by applying the following idea to consider of them:

Contradistinction of brightness. Often we feel a color brighter if it is surrounded by dark colors. To apply this mechanism, our implementation introduces the variable b for bright colors. If the bright color occupies small parts of an image, and it is surrounded by larger parts of dark colors, our implementation applies the value $b > 1$ to the pixels of the bright colors. Otherwise, it applies the value $b = 1$ for each pixel.

forward and backward movement. Often we feel forward movement for warm colors, and backward movement for cool colors. To apply this mechanism, our implementation introduces the variable p for each of 66 colors, which applies the value $p > 1$ for warm colors, and $p = 1$ for cool colors.

Applying the above two variables, our implementation calculates C_i using the following equation:

$$C_i = \sum_{j=1}^N b_j p_j f_j \quad (1)$$

Here, N is the total number of pixels in the image, b_i and p_i are the values of b and p at the j -th pixel, and f_j is a binary value, where $f_j = 1$ if j -th pixel is i -th color, otherwise $f_j = 0$.

After calculating C_i , our implementation leads the following equations to calculate association of rhythm patterns from an input image:

$$P_j = a \sum_{i=1}^{66} C_i R_{ij} + b \sum_{k=1}^M Q_{kj} \quad (2)$$

Here, P_j denotes the association of j -th rhythm pattern, a and b are constant positive values, R_{ij} is the association of i -th color to the j -th rhythm pattern, M is the number of keywords, and Q_{kj} is the association of k -th keyword to the j -th rhythm pattern.

The preparation stage obtains the values of R_{ij} and Q_{kj} by the combination of the following procedures.

Procedure 1. DIVA shows a list of keywords to users, and collects the feedback of values of Q_{kj} for each combination of keyword and rhythm pattern.

Procedure 2. DIVA shows the 66 colors to users, and collect the feedback of values of R_{ij} for each combination of color and rhythm pattern.

Procedure 3. DIVA shows many sample images (25 images in our experiments) to users, and collect the feedback of values of P_{jl} , the association between j -th rhythm pattern and l -th sample image. Then it obtains optimal R_{ij} values, by applying an optimization scheme to the following equation:

$$P_{jl} = \sum_{i=1}^{66} C_{il} R_{ij} \quad (3)$$

Here, C_{il} is the importance of i -th color in the l -th sample image.

The actual stage calculates P_j values of the all rhythm patterns for an input image, using equation 2 with the values of R_{ij} and Q_{kj} obtained by the preparation stage. It then selects the rhythm pattern that obtained the highest P_j value.

Finally, DIVA arranges the input music applying the selected rhythm pattern. We assume that input SMF consists of four parts, melody, harmony, bass, and percussions. Our current implementation simply replaces bass and percussions parts by the selected rhythm pattern, without applying any modifications for melody and harmony parts.

3 Experiments

This section shows our experiments with DIVA, preparing 7 rhythm patterns, 10 keywords, and 25 sample images. The keywords include the names of objects shot in the images, such as sunset, flower, and snow. We prepared four types of preparation stages as follows.

Preparation 1. We applied procedure 1 to obtain Q_{kj} , and let $R_{ij} = 0$.

Preparation 2. We applied procedure 2 to obtain R_{ij} , and let $Q_{kj} = 0$.

Preparation 3. We applied procedure 3 to obtain R_{ij} , and let $Q_{kj} = 0$.

Preparation 4. We applied procedure 3 to obtain R_{ij} , and procedure 1 to obtain Q_{kj} .

For the evaluation of automatic arrangement results, we provided all 7 arrangements to users, and collected the selection of arrangement associated from the input images. After that, we applied DIVA to the input images, and calculated the ratio of concordance between users' and DIVA's choices of rhythm patterns.

We asked the above procedures to 8 experimental users, including 3 male and 5 female, 3 workers and 5 students, 5 experts and 3 non-experts in music. Ages of all experimental users were 20's.

Table 1. Ratio of concordance between users' and DIVA's choices.

Preparation 1	Preparation 2	Preparation 3	Preparation 4
0.11	0.28	0.44	0.71

The result shows that DIVA provided good results, since random selection will bring 0.14 of ratio of concordance in average, and the above results were much better than the random selection.

This result also gives some suggestions: The ratio of concordance is better when we apply both color distribution and keywords for the calculation of association of rhythm patterns. Also, it is better to provide sample images rather than predefined colors to users.

We observed that experimental users can be divided into the following two patterns:

- Some users present higher association between colors and rhythm patterns.
- Others present higher association between keywords and rhythm patterns.

We modified the constant values as $a > b$ in equation (2) for former users, and $a < b$ for latter users, and observed that the ratio of concordance got higher.

4 Conclusion

This poster presented a technique for automatic music arrangement, which selects the rhythm pattern most associated from color distribution and keywords of input images. A part of our experimental results are shown at http://itolab.is.ocha.ac.jp/kisa/diva1/index_e.html.

Our future works include: larger experimental tests with more sample images and more rhythm patterns, more detailed analysis of feedback of experimental users, more sophisticated implementation of color contradistinction, and more sophisticated calculation of R_{ij} values in procedure 3.